

The 20th Century Reanalysis Project

Gilbert P. Compo^{1,2}, Jeffrey S. Whitaker², and Prashant D. Sardeshmukh^{1,2}

¹University of Colorado/CIRES Climate Diagnostics Center, USA

²NOAA Earth System Research Laboratory/Physical Sciences Division, USA

Long-term climate datasets are critical to understand the causes of climate variability, to assess its potential predictability, and to evaluate its simulation in climate models. New climate datasets are needed to evaluate the weather and climate variability of the climate model simulations used in the Intergovernmental Panel on Climate Change (IPCC) Assessment Report 4 (AR4) and to contribute to model improvement for subsequent reports. Over the past decade, major national and international efforts have led to the creation of the first generation of climate datasets called “reanalyses,” which for the dataset jointly created by the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR, Kalnay et al. 1996) spans the period of substantial upper-air observations (1948-present). Prior to this period, the only long-term atmospheric reanalysis datasets consist of Northern Hemisphere sea-level pressure maps that were hand drawn in the 1940s from available, but incomplete, surface records (United States Weather Bureau, 1944). The US Climate Change Science Program (CCSP) Strategic Plan (Mahoney et al. 2003) strongly emphasizes the importance of developing new reanalysis data sets to improve representations of past climate variability and change, identifying “Reanalyses of historical climate data for key atmospheric features” as one of 21 high-priority products. The Global Climate Observing System Implementation Plan (IP, Mason et al. 2004), a component of the Global Earth Observation System of Systems, is emphatic in urging parties “to develop improved methods for such reanalysis” (IP, p. 10). While the current reanalysis datasets have been valuable for climate research and applications, the fact that they extend back to only the mid-20th century greatly limits their usefulness for addressing numerous science questions of high societal relevance, such as the major features of the atmospheric circulation during the 1930’s U.S. Dust Bowl and determining long-term trends in the frequency and severity of meteorological phenomena such as severe storms, floods, cold spells and heat waves. The quality of the AR4 simulations in representing these and other meteorological phenomena with socio-economic impacts is an important issue as decision makers respond to the recent IPCC report. To meet these needs, the 20th Century Reanalysis Project is using improved observational data sets together with state-of-the art data assimilation methods to create a 6-hourly, 3-dimensional atmospheric reanalysis dataset spanning 1892 to 2007.

Over the past several years, we have developed a capability to produce high-quality six-hourly reanalyses for the troposphere from surface pressure observations alone (Compo et al. 2006) using a data assimilation system based on the Ensemble Kalman Filter (Whitaker and Hamill 2002). Several recent studies have established the feasibility of producing a reanalysis dataset from the 1890s to present (Whitaker et al. 2004, Anderson et al. 2005, Compo et al. 2006). Encouraged by these results, we have begun producing a century-long reanalysis dataset. Initial results span 1918-1958.

To produce the 20th Century Reanalysis dataset, we have used the Ensemble Filter system as described in Compo et al. (2006). The Filter has a simple, but parallelizable, implementation when observations are processed one at a time. Assume that we have an n -member ensemble of first guess fields, with the j th member \mathbf{x}_j^b representing the complete state vector of the forecast model (e.g., wind, temperature, humidity, and surface pressure fields on the model domain). We have chosen $n=56$ based on tests showing that ensembles smaller than 50 members degrade the quality of the upper-air analyses, and the fact that 56 conveniently corresponds to the CPU configuration available for our implementation of the

Ensemble Filter. The sample mean of these fields is $\bar{\mathbf{x}}^b = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^b$, and the deviations from the mean are

$\mathbf{x}'_j{}^b = \mathbf{x}_j^b - \bar{\mathbf{x}}^b$. We denote the first surface pressure observation to be assimilated as y^o and the ensemble

mean and deviations interpolated to the observation location as $\overline{y}_j^b = \mathbf{H}\overline{\mathbf{x}}_j^b$ and $y_j'^b = \mathbf{H}\mathbf{x}_j'^b$, respectively, with \mathbf{H} being the operator that interpolates the first guess surface pressure field to the observation location. We combine the first guess ensemble and the observation to form an n -member analysis ensemble, whose mean $\overline{\mathbf{x}}^a$ and deviations $\mathbf{x}_j'^a$ are calculated via

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^b + \mathbf{K} \left(y^o - \overline{y}^b \right) \quad (1)$$

and

$$\mathbf{x}_j'^a = \mathbf{x}_j'^b - \tilde{\mathbf{K}} \left(y_j'^b \right) \quad (2)$$

where the Kalman gain \mathbf{K} is given by

$$\mathbf{K} = \frac{1}{n-1} \sum_{j=1}^n \mathbf{x}_j'^b y_j'^b \left(\frac{1}{n-1} \sum_{j=1}^n y_j'^b y_j'^b + R \right)^{-1}, \quad (3)$$

and the modified Kalman gain $\tilde{\mathbf{K}}$ is given by

$$\tilde{\mathbf{K}} = \left(1 + \frac{R}{\frac{1}{n-1} \sum_{j=1}^n y_j'^b y_j'^b + R} \right)^{-1} \mathbf{K}. \quad (4)$$

We assume that the observation y^o has a specified observational error variance R that represents both the measurement error associated with the observation and the error associated with representing a large area (i.e., a numerical weather prediction model grid box) from a point measurement. We further assume that the error in y^o is uncorrelated with all the other observations to be assimilated. The model means, variances, and co-variances in Eqs 1-4 are all unbiased sample estimates from the $n=56$ member ensemble. Within the limitations of using an imperfect model and finite ensembles, this formulation represents a minimum-error estimate of the “true” state (Lorenc 1986), represented here by the analysis ensemble mean $\overline{\mathbf{x}}^a$. The uncertainty for this estimate is given by the covariance of the analysis ensemble deviations $\mathbf{x}_j'^a$. To assimilate subsequent observations, the 56 members of the analysis ensemble \mathbf{x}_j^a become the new first guess ensemble and equations 1-4 are applied iteratively for each observation. After all available observations have been assimilated, the 56 member set of analyses $\mathbf{x}_j^a = \overline{\mathbf{x}}^a + \mathbf{x}_j'^a$ becomes the 56 initial conditions for the subsequent 9-hour forecast/analysis cycle (see Whitaker et. al. 2004 for a complete description). In a significant modification to the system of Compo et al., hourly observations are used in a 6-hour window centered on the analysis time. Equations 1-4 remain unchanged, but now the covariance in (3) represents time dependent information between a $y_j'^b$ at a possibly different time than $\mathbf{x}_j'^b$. This allows information from fast-moving weather systems to be used to improve the analysis, letting the analysis “know” where the system will be in the future.

The short-term forecast ensemble is generated in parallel from 56 9-hour integrations of a state-of-the-art atmospheric general circulation model, the atmospheric component of NCEP’s operational Climate Forecast System model (Saha et al. 2006). Briefly, the model has a spatial resolution of nearly 200-km on an irregular Gaussian grid in the horizontal (corresponding to a spherical harmonic representation of model fields truncated at total wavenumber 62, T62). In the vertical, for efficiency, we have reduced the resolution from a finite differencing of 64 levels to 28 levels with no detrimental effects to the reanalysis quality. The model top is at 0.2 hPa. The model has a complete suite of physical parameterizations as described in Kanamitsu et al. (1991) with recent updates detailed in Moorthi et al. (2001). Additional

updates to these parameterizations, specific to this version of the model, are described in Saha et al. and include revised solar radiation transfer, boundary layer vertical diffusion, cumulus convection, and gravity wave drag parameterizations. In addition, the cloud liquid water is a prognostic quantity with a simple cloud microphysics parameterization. The radiation interacts with a fractional cloud cover that is diagnostically determined by the predicted cloud liquid water. The specified boundary conditions needed to run the model in atmosphere-only mode are taken from the time-evolving sea surface temperature and sea ice fields of the HadISST1.1 dataset obtained courtesy of the United Kingdom Met Office Hadley Centre (Rayner et al. 2003).

Figure 1 illustrates the Ensemble Filter capability using Eqs 1-4 and the CFS atmospheric model. It shows the degree to which the principal mid-tropospheric features for an extreme event, the famous start of the Battle of the Bulge of 16 December 1944 are present in a map from our reanalysis using surface pressure observations alone and the Ensemble Filter (right panel, $n=56$). We have compared our reanalysis field to the features seen in a map which used all available surface, radiosonde and other upper-tropospheric observations: a hand-drawn real-time map from the Air Weather Service (left panel). It is remarkable that our reanalysis, using only the surface pressure observations and a relatively low-resolution model, is able to replicate many of the features seen in the hand-drawn mid-tropospheric analysis produced at the time. Most likely, this ability arises from our use of the Kalman gain in Eqs 1 and 2 which adjusts for both the meteorological conditions and the observational network. Overall, our feasibility studies suggest that the extratropical, upper-tropospheric Northern Hemispheric height errors obtained from Ensemble Filter-based analyses that use even fewer surface pressure observations will be comparable to current 2-3 day weather prediction forecast errors (e.g., 500 hPa geopotential height spatial correlations of ~ 0.95 , root-mean-square errors ~ 35 -40m).

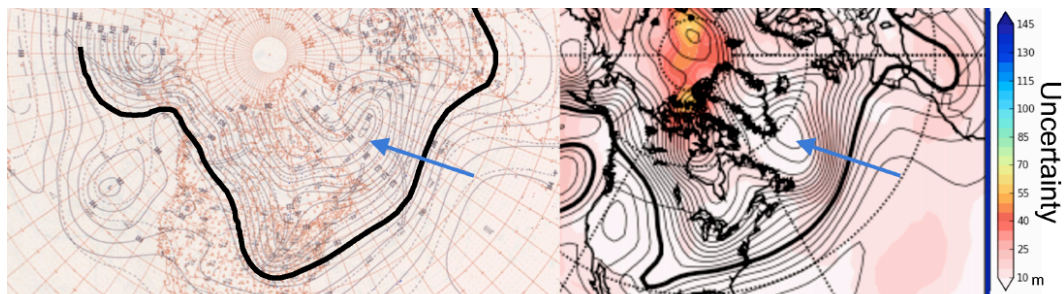


Fig. 1: The 500 hPa geopotential height analysis of 16 December 1944 06GMT from the Ensemble Filter data assimilation system using only surface pressure observations (right) and an Air Weather Service map drawn for 04GMT in near-real time is also shown (left). Colored arrows illustrate the same features in both maps. Note that the left panels used all available surface, radiosonde and other upper-air observations. Contour interval is 200 ft (50 m), and the 18000 ft=5608 m (5600 m) line contour is thickened, in the left (right) panel. Shading in the Ensemble Filter analysis indicates the analysis uncertainty.

Figure 2 provides additional evidence that the Ensemble Filter-produced upper-level tropospheric circulation fields will reflect the actual atmospheric variations. Shown in the black dots are anomalies of newly digitized radiosonde observations of 500 hPa (approximately 5500 m altitude) and 300 hPa geopotential height at Ilmala, Finland (62.2°N, 24.92°E) for the period 1935 – 1956 (expanded from Bronnimann 2003) compared to analyzed anomalies from the Ensemble Filter. The upper-air variability produced by the reanalysis at this high latitude location appears to be consistent with the direct measurements, even on a case-by-case basis. The fidelity of the comparison suggests that the Ensemble Filter will be able to reconstruct upper-air variability of both weather and climate variations throughout the 20th century.

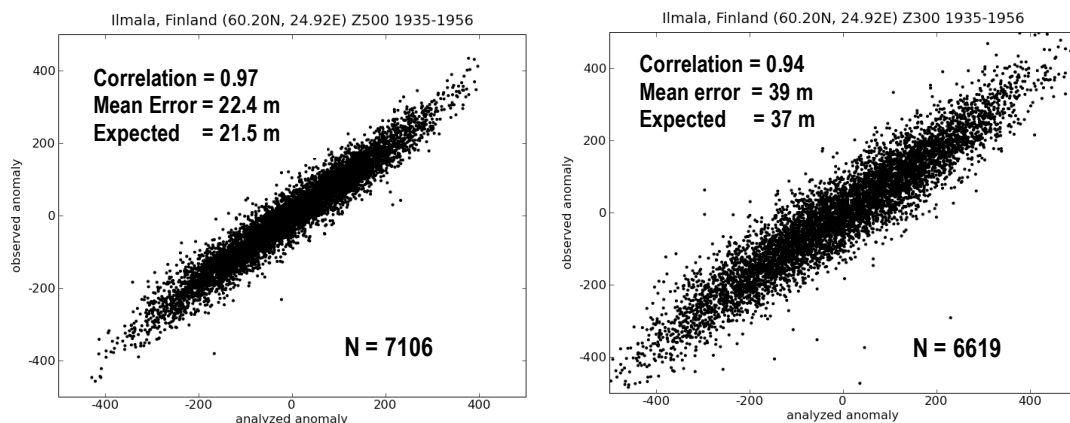


Fig. 2: Observed anomalies of 500 hPa (left) and 300 hPa (right) geopotential height from radiosonde measurements taken at Ilmala, Finland over the period 1935-1956 compared with analyzed anomalies from the Ensemble Filter using only surface pressure observations. The Ensemble Filter estimates are interpolated to the temporal and spatial coordinates of the Ilmala data. The Ensemble Filter is able to represent most of the variability seen in the observations with correlation coefficients of (left) 0.97 and (right) 0.94 (Ilmala data courtesy of A. Grant and S. Bronnimann of ETH-Zurich).

The combination of this data assimilation method with a new database of station and sea level pressure observations provides an opportunity to produce, for the first time, a reanalysis data set of a century or longer. Under the auspices of the GCOS AOPC/OOPC Working Group on Surface Pressure and GCOS/WCRP Working Group on Observational Data Sets for Reanalysis, and in collaboration with the Atmospheric Circulation Reconstruction over the Earth (ACRE) Initiative led by Dr. Rob Allan of the UK Hadley Centre, NCAR, CIRES, and NOAA's Earth System Research Laboratory and National Climatic Data Center have combined international collections of land and marine surface and sea level pressure observations to create the most complete data set of pressure observations ever assembled, called the International Surface Pressure Databank. For 1892 to present, with our national and international partners, we have extracted the complete set of ICOADS 2.1 and ICOADS Auxiliary marine sea level pressure observations (Worley et al. 2005) and combined all available surface and sea level pressure subdaily land observations. Partners contributing to the ISPD include the All Union Research Institute Hydrometeorological Information WDC; Atmospheric Reconstructions over the Earth; Australian Bureau of Meteorology; British Antarctic Survey; Danish Meteorological Institute; Deutscher Wetterdienst; EMULATE; Environment Canada; ETH-Zurich; GCOS AOPC/OOPC Working Group on Surface Pressure; GCOS/WCRP Working Group on Observational Data Sets for Reanalysis; Hong Kong Observatory; ICOADS; Instituto Geofisico da Universidade do Porto; Japanese Meteorological Agency; Jersey Met Dept.; KNMI; MeteoFrance; Meteorological and Hydrological Service, Croatia; NOAA Earth System Research Laboratory, National Climatic Data Center, National Centers for Environmental Prediction, Northeast Regional Climate Center at Cornell U., Midwest Regional Climate Center at UIUC; Norwegian Meteorological Institute; Ohio State University, Byrd Polar Research Center; Proudman Oceanographic Laboratory; SIGN - Signatures of environmental change in the observations of the Geophysical Institutes; South African Weather Service; UK Met Office Hadley Centre; U. of Colorado-CIRES/Climate Diagnostics Center; U. of East Anglia-Climatic Research Unit; U. of Lisbon-Instituto Geofisico do Infante D. Luiz; U. of Lisbon- Instituto de Meteorologia; U. of Milan-IFGA; and U. Rovira i Virgili-CCRG.

In this project, we are capitalizing on these advances and have begun producing a complete 6-hourly, 3-dimensional reanalysis dataset from the beginning of the first complete continental United States meteorological observing system (1892) to the present. The current 1918-1958 dataset will be released in Spring 2008. The complete dataset will be produced by Spring 2009. This new dataset will extend the

record of tropospheric gridded fields 56 years before the start of the NCEP-NCAR reanalysis, back to a period for which no gridded upper-tropospheric analyses are available, and will also overlap for comparison with it and other reanalysis datasets. Production of the dataset, together with the *quantified uncertainties* produced by the assimilation system (the covariances of the ensemble member deviations in Eq. 2, e.g., shaded regions in Fig. 1), will provide the scientific community with the information necessary to meet U.S. and international needs.

REFERENCES:

- Anderson, J. L., B. Wyman, S. Zhang, and T. Hoar, 2005: Assimilation of surface pressure observations using an ensemble filter in an idealized global atmospheric prediction system. *J. Atmos. Sci.*, **62**, 2925-2938.
- Bronnimann, S., 2003: A historical upper air-data set for the 1939-44 period. *Int. J. Climatol.*, **23**, 769-791.
- Compo, G.P., J.S. Whitaker, and P.D. Sardeshmukh, 2006: Feasibility of a 100-year reanalysis using only surface pressure data. *Bull. Amer. Met. Soc.*, **87**, 175-190.
- Group on Earth Observations, 2005: The Global Earth Observation System of Systems 10-year Implementation Plan, [Available online at <http://earthobservations.org/docs/10-YearImplementationPlan.pdf>]
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Met. Soc.*, **77**, 437-471.
- Kanamitsu, M, and Coauthors 1991: Recent changes implemented into the global forecast system at NMC. *Wea. Forecasting*, **6**, 425-435.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177-1194.
- Mahoney and Coauthors, 2003: Strategic Plan for the U.S. Climate Change Science Program. Report by the Climate Change Science Program and the Subcommittee on Global Change Research [available online at <http://climatescience.gov/Library/stratplan2003/default.htm>]
- Mason, P.J., and Coauthors, 2004: Implementation plan for the global observing system for climate in support of the UNFCCC. WMO-IOC-UNEP-ICS, GCOS-92, Technical Document No. 1219, World Meteorological Organization, Geneva, Switzerland, 153pp.
- Moorthi, S., H.-L. Pan, and P. Caplan, 2001: Changes to the 2001 NCEP operational MRF/AVN global analysis/forecast system. NWS Tech. Procedures Bulletin 484, 14 pp. [Available online at <http://www.nws.noaa.gov/om/tpb/484.htm>.]
- Rayner, N.A., D.E. Parker, E.B. Horton, C.K. Folland, L.V. Alexander, D.P. Rowell, E.C Kent, and A. Kaplan, 2003: Global analyses of SST, sea ice and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Saha, S. and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483-3517.
- United States Weather Bureau, 1944: *Daily Synoptic Series, Historical Weather Maps, Northern Hemisphere Sea-Level, January 1899 to June 1939*. Cooperative project of U.S. Army, Air Force, and Weather Bureau Statistics Division.
- Whitaker, J. S., and T.M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913-1924.
- Whitaker, J. S., G.P. Compo, X. Wei and T.M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190-1200.
- Worley, S.J., S.D. Woodruff, R.W. Reynolds, S.J. Lubker, and N. Lott, 2005: ICOADS Release 2.1 data and products, *Int. J. Climatol.*, **25**, 823-842.